

論 説

心理学研究の技法－論文読みから実験の計画まで (I)

小牧 純爾¹⁾

An introduction to strategies and tactics for research in Psychology: (1) How to read journal articles of experimental studies

Junji KOMAKI¹⁾

はじめに

心理学の研究は、「読む」「書く」「実証する」の3つの局面に分けて考えることができる。心理学の研究者は、人によってウェイトに差はあるが、論文や専門書を読み、論文や本を書き、実験や調査を実施する。これらの局面には、それぞれに、外してはいけなポイントやルールがあり、的確に進めるための「工夫」や「秘訣」がある。たとえば、実験論文を読む時には、「結果」から切り離して「方法」を読むべきである。実験の結果を主体的な枠組みで検討するためである。あるいは、実験を実施する時には、データの分析法を配慮して計画を策定するべきである。データが得られてから適切な統計処理の方法がないことに狼狽するようなことを避けるためであり、研究の焦点を確認しておくためでもある。この論説ではこうした工夫や秘訣などについて、実験研究を中心に意見を述べたい。分量の関係もあり、第1章「論文を読む」、第2章「実験を実施する」、第3章「論文を書く」に分けて発表することにした。

第1章 論文を読む²⁾

論文読みの必要性

論文読みは研究の要件 心理学の研究者にとって、心理学の専門論文を読むことは欠かすことのできない研究の一環である。研究と教育にたずさわる教員なら、日常的に専門の雑誌には目を通す。そして、自分

の研究領域の論文はもとより、領域外のものであっても、新しい研究報告は注意深く検討し、知識を更新して将来に備えるのが普通である。教員に比べ渉猟の範囲は狭いものの、大学院生の場合も事情は同じである。自分の研究テーマに関係する論文がないか、内外の主立った専門雑誌には目を通す。加えて、各種のデータベースを検索し、重要な研究の見逃しがないように努力する。

学部の専門コースで心理学を専攻する学生など、心理学の初学者も例外ではない。ゼミナールや文献購読の授業では、専門の論文を読み、内容について報告し、討論することが必修になっている。また、卒業論文や修士論文を書く時には、関連する文献を読み、考察を加えることが期待されている。学部の段階では専門論文の検証を求めている大学もある。しかし、修士論文や博士論文では、関連論文を中心とする文献の検討なしには、まず執筆は不可能である。それに、文献研究を欠いた、いわゆるモノローグでは論文審査も安泰ではない。

専門論文の特殊性 文献購読やゼミナールで初めて専門論文を読むことになった時、初学者のほとんどは途方に暮れるだろう。専門論文は日頃お馴染みの文書とは肌合いが違う。まず、使われている用語に馴染みがない。論文では、たとえば「表象連合」とか「隠蔽」とか非日常的な言葉がいきなり出てくる。おまけに、論文の各ページには文字ばかりがベタ一面に並んでおり、内容の見当をつけるためのイラストレーションなどはまず見あたらない。

専門論文のこうした取り付きにくさは、専門論文が対象を研究者に限定した、特殊な文書であることに理由がある。論文は、研究者が研究の目的、方法、結果、それに考察を他の研究者に伝えるための文書であり、何よりも正確で明瞭であることが重視される。このこ

¹⁾ 金沢大学名誉教授

Kanazawa University

²⁾：統計処理パッケージや「効果量」の利用の現状については、金沢大学の谷内通氏から教示をうけた。また、本論説の草稿に対し、谷内氏のほかに、2名の査読者から有益なコメントと助言をいただいた。感謝いたします。

とから、論文では、ほとんどの場合、多義的で意味のあいまいな日常用語は使われない。それに代わり、研究者間で定義が共有されている専門用語が用いられる。また、論文では、研究の核心を簡潔に、かつ、正確に表明することが不文律になっている。余見の恐れのある刺激的な用語や強調表現は論文の文章には不適當であるとされている。それに、表現の冗長さを排除するところから、素人を意識した丁寧な説明は必要ないと考えられている。

さらに、専門論文には、初学者の意欲をそぐ顕著な特徴がある。それは内容の非現実性である。論文の研究はいかにも学術的で堅牢である。しかし、その論証は人工的で非日常的な方法を用いていることが多く、示されている結論も、ほとんどの場合、一般的で抽象的である。論文が日常的な問題の解決のためにどのような展望を与えてくれるのか、素人には釈然としない場合が多い。この問題は実証研究の方法論の重要なポイントである。後の章で改めて論じることにした。

専門論文のリテラシー

論文の公共性 同じ論文を読んでいても研究者によって理解に差があることがある。情報の受容には、一般に、個人の関心、知識などが深く関わっており、人によって文書の読み取りに違いがあるのは周知のことである。しかし、そのことがわれわれの日常で問題になることは少ない。ちなみに、小説を楽しむといった趣味の世界では、読みの違いに目くじらを立てる人はいない。小説は読み手の感興に合わせ、自在に読んでいいのである。しかし、論文読みでは話はまったく違う。論文を読むのは趣味の世界の個人的な行為ではなく、研究交流という公共的な活動の出発点となる行為である。公共的な研究の世界では、個人の恣意性は許されない。

論文には、読み飛ばしてはならないポイントがある。極端な事例になるが、論文の「要約」を読んだだけでは論文を読んだことにはならない。これは本の目次に目を通して本全体を読んだことにすると類似であり、論外である。また、いうまでもないが、論文の内容に一面的な解釈を加えてはならない。論文には、読んでも理解できない個所が出てくることが多い。しかし、そうした場合には、新たな文献調査を行うなど、別の手立てを講じ、正しい理解に達することが必要である。ポイントの見逃しや内容の無意識的、あるいは、意識的誤読は、議論に無用の混乱をもたらすだけである。

論文のポイントを予断なく、正しく読み取ることで始めて公正な研究交流がスタートする。このことは論文に対する疑問や意見を差し控えることを意味しない。疑問や意見はきちんと論文を読んだ後で「それと分かる形」で提起するべきであり、論文の読み取りの段階に読み手の理論的立場などを混入するべきではないのである。論文を適正に読み取った後での疑問や意見の表明は、むしろ、研究交流の健全性の表れであるといつてよい。

論文読みの準備 専門の論文を読むとして、いきなり論文に取りかかるのは得策ではない。まず参考文献に目を通し、論文の扱っている問題について予備知識を身につけた方が、論文読みはうまく行く。参考文献のリストを活用したい。ただし、然るべき参考文献で「目次」や「事項索引」に当たっても、適当な解説が見つからないことがある。新しい研究テーマの場合には、特にそうであり、まず見つからないと覚悟した方がよい。これには、研究が進行中で大方の理解が一定していない問題については、異論のない解説を書くのがむずかしい事情が関係している。こんな場合には、とりあえず、研究テーマの背景にある研究史を知っておくのがよい。それだけでも論文読みの相当の助けになる。

論文を読む時には、関連領域の事典や辞典類は手許に確保したい。論文を読み始めて未知の用語に出くわした時、定義や意味を確かめるためである。なお、英語で書かれた専門文献を読む場合には、用例が数多く載っている辞書を用意したい。記載されている訳語候補の中から原文の文脈に合った用例を見つけ、その用例の訳語を選定することができるからである。この、用例に焦点を絞って訳語を選定する方法は、英和辞典よりも英々辞典の方がやりやすい。英々辞典の語彙項目には、当然のことながら、日本語の訳語ではなく、語の「説明」が記載されている。このことから、記載されている用例と説明とを手がかりに、自分の知っている日本語の語彙の中から、検索語に対応する日本語を選び出すことができる。英々辞典を使う時には、したがって、英文の文脈を読み取る能力に加え、確かな語彙を選び出す「日本語の能力」が必要になる。

30 年ほど昔、学習理論に関する専門書を翻訳中に、“formulate”という動詞の訳出にてこずったことがある。当時使用していた英和辞典では、第 1 義に「公式化する」「公式で表す」、第 2 義に「明確に系統立てて説く(計画する)」といった訳語が出ていた。しかし、

原文は別に数学の公式について論じてはいなかったし、何かの説得や計画について解説してもいなかった。困惑のあげく、ある現代英々辞典を参照したところ、第1義の用例の一つ

Charles Darwin formulated the theory of natural selection.

という用例に添えて「アイデアまたは理論を形成する」という説明が記載されていた。そこで、私は英和辞典にはなかった「理論化する」という訳語を“formulate”に当てた。そして、関連する“formulation(s)”という可算名詞を「(諸)理論」と訳すことにした。辞典にない訳語を選ぶには勇気がある。しかし、大事なことは文脈に添って正しく文意を読み取ることである。このことを重視したい。なお、辞書には適不適がある。現在の英語専門論文を読むには、現代米語にも対応した「現代英々辞典」を選びたい。

論文の構成と読みの順序

論文の構成 実験論文は大体決まった構成で書かれている。冒頭にくるのは論文全体の内容を短く集約した「要約」abstractで、最近では日本語の専門誌論文でも、英語で書かれていることが多い。次に研究が何のために行なわれたのか、研究の目的を述べた「目的」purposeが来る。そして、その後に「方法」method、「結果」results、「考察」discussion、「参考文献」referencesと続く。方法はどんな被験体(実験参加者)や装置や手続きで実験を行ったのかの説明であり、結果は用いた方法のもとでどんなデータが得られたのかに関する、結果と結果分析の報告である。考察は得られた結果に関する著者の見解を提示する論文の核心部分であり、目的が達成されたのか、他の研究との関連性はどうなのかなどが主要な論点になっている。参考文献は参照した情報源を表示するリストである。

読み始めは「方法」から 論文を読む順序は重要である。多くの場合、要約と目的に軽く目を通した後、方法から読み始めるのがよいと私は考えている。知識がまだ十分でない初学者は特にそうした方がよい。その理由の第一は、方法が、被験体(実験参加者)、実験装置、手続きなど、方法を構成している実験変数 variables を具体的に説明しているセクションであり、抽象的でない分、取りつきやすいからである。方法から読み始めることには、ほかにも重要な2つの理由がある。それは、結果より先に方法を読むことで、結果

から切り離して、実験方法の「妥当性」を検討できるようにするためであり、さらには、結果を見る前に、結果の出方を「推理」できるようにするためである。

条件の設定や手続きなどの実験の方法は、基本的には、「こうすれば」「こうなる」という法則的知識に支えられて決定されている(小牧, 2000, p.105)。つまり、こうした方法を用いれば(「こうすれば」), 狙いの心的過程を実験的に生じさせることができる(「こうなる」)といった知識にもとづいて、方法が定められているのである。ただし、この法則的知識は固定的ではなく、研究の進歩にしたがって修正され、更新されることを理解しておくことが重要である。実験の方法は新たな知見にしたがって改良され、それに伴って方法の妥当性に関する知識も更新されてくるのである。

すべての実験方法に妥当性がないと主張するつもりはない。実験の方法が現在も蓄積されつつある研究の成果を適正に反映しているかどうか、それを慎重に検討することが必要だと指摘しているのである。そして、さらに一步を進め、方法を結果から切り離して読むことで、実験で用いられている方法が適切かどうかを「結果の先入観」なしに検討する機会を設けることを勧めているのである。狙った結果が出ているかどうかを先に確かめると、結果に引きずられて妥当性の吟味がおろそかになってしまうおそれがある。このことは、研究が始まったばかりで、実験方法の妥当性がまだ十分に確立されていない領域の論文を読む時に、特に重要となる。

「結果」を推理してみよう 方法を読み終えたとして、結果の読みに取りかかるまえに、方法に書かれていた情報だけをもとに、頭の中で仮想実験をやってみることを勧めたい。つまり、読んだばかりの方法を情報源にして、どんなデータが得られることになるのか推理してみるのものである。新たな実験を計画したことのある研究者なら、結果が出る前にその実験の結果がどうなるのか、予測した経験があるはずである。これと同じことを読んでいる実験についてもやってみるのである。

実験の結果がすぐ後のページに示されているのに、なぜに回り道になるようなことを行うのか。それは報告される実験の結果を適正に理解し、評価するための確かな「枠組み」を確立するためである。まず、予測を立ててみることで、方法の読み取りに齟齬がなかったか、もう一度確かめることができる。方法に見落としがあったり、誤解があったりすると、結果を予測す

ることがむずかしくなる。また、予測を立ててみることで、方法の説明に重大な記載もれがあることに気がつくこともある。さらに重要なことは、方法のもたらす効果、先の表現では「こうなる」を自分なりに予測してみることで、実験者の採用した方法が適切であったかどうかを、結果に引きずられることなく、客観的に検討することができるからである。

実験の計画に、その種の実験には用いられたことのない新しい操作が導入されている場合には、事前の推理はむずかしい。そんな場合には未知の操作の効果に幅を持たせた予測を立てることが考えられる。さて、キャリアのある研究者なら、論文の結果を読んだ際、反射的に「なるほど」とか「なにかおかしいぞ」とかを感じた経験があるはずである。これは、その研究者が、論文の結果の読みにとりかかる前から、それと意識しないままに結果を予測していた証拠である。「なるほど」と感じた時は、研究者の持っていた法則的知識に、論文に示された結果がおおむね整合的であることが示唆される。一方、「なにかおかしいぞ」と感じた時には、研究者の予測にそわない結果が示されたことを意味する。知識の乏しい初学者には、こうした事前予測を立てることはむずかしい。しかし、特定のテーマを扱った論文を読んで行くうちに、そのテーマの論文については、じきにこうした予測を立てることができるようになってくる。

実験結果を点検する

実験結果の統計的信頼性 「結果」のセクションには、得られたデータとその分析結果が、たいいていの場合、図や表を添えて提示されている。結果を読み進める主要なポイントは、こうしたデータとその分析結果が、意図した目的に対する確かな答えになっているかどうかを判定することにある。

こうした判定を進めて行く最初の点検ポイントは、実験データの統計的な信頼性である。つまり、報告されているデータが、偶然に得られた不確かな素性のデータではなく、一般性につながる推論を導くための確かな証拠であると判断してよいかどうかという問題である。これは内容とは別の、データ一般に関わる問題である。この信頼性の判定には、これまでは、データを推測統計法にかけて処理し、「統計的有意性」を確認することによって行われてきた。このことから、結果読みの最初の課題は、統計的に有意な結果が得られているかどうかを確認することにあった。ただし、最

近では、こうした有意性の検定に加えて、要因作用の「効果量」や「信頼区間」に関する統計的推定が求められるようになってきている。こうした統計的推定法を用いる信頼性の判断は、客観的で公共的であることを強調しておきたい。有意性や効果量や信頼区間の統計的検定は方法が明示されている。適切な方法を選定し、正しく演算を行えば、初学者であろうと、キャリアの豊富な研究者であろうと、誰がやっても同じ結果が得られる。これは判定の公共性を保証する重要なことだからである³⁾

さて、データの統計的信頼性を点検する課題とは、具体的には、適切な検定法が用いられているか、そして、適切なモデルにしたがって正しく計算が行われているかを検討することである。最近では誤った検定法を適用する初歩的な誤りは少なくなっているが、 F テストのモデル選定などには、現在でも、問題のあるケースが認められる。よく指摘される事例は、記憶実験などで材料の「語」を固定要因として扱う場合である。語をランダム要因ではなく、固定(確定) 要因とするモデルでは、語と他の要因との交互作用分散の期待値がゼロとして処理されてしまう(小牧, 1995, p.37-45)。このことから、語を固定要因とするモデルでは、検討する変動因の F 値が、実際には、モデルの想定外の成分を含んだ値にもとづいて、計算されることになる。この結果、検討する語関連要因が「不当」に有意な値をとる可能性が出てくるのである(小牧, 2000, p.167-171)。こうした不適切なモデルを採用した有意性検定は「留保条件」をつけて読み取る必要があると思われる。

一方、検定法や検定のモデルについて、論文に十分な説明がない場合がある。そんな時には、自由度 df などの検定法のパラメータを参照し、検定法やモデルを推定することが必要になる。たとえば、ラットの正選択率がチャンス水準以上であることを検定するために、 t テストを用いることがあるが、この場合には2つの平均値を比較する2標本 t テストではなく、「1標本 t テスト」one-sample t test が正しい検定法になる。この検定が用いられている場合、自由度は標本の大きさ

3) : こうした統計的推定法を用いるアプローチにくみしない立場がある。個体型実験法を用いる研究者は、「目による検査」visual inspection によって要因の効果を判定することが多い。また、効果の一般性、ここでの効果の信頼性については、研究者の研究実績に裏付けられた経験的基準によって判定することを認めている。しかし、実験結果の信頼性の判定が研究者の研究経験にもとづいて行われるのであれば、これは一種の権威主義であり、問題であると私は考えている(小牧, 1992, p.140)。

から 1 を引いた値 ($n - 1$) になる。ほかにも、多元要因の F テストで、何を誤差項にして要因の効果を計算したのかが不明であることがある。こうした時には、方法に記載されているそれぞれの要因の水準数を参照し、 F 値に添えられている分母の自由度 df_2 を見た上で、誤差項を推定する。平均誤差平方 MSE を含め、こうしたパラメータの記載のない論文は、有意性の証明されなかった実験の場合と同じに扱うほかはない。

ただし、検定結果が統計的に有意でないとして、論文に「学問的価値」がないと結論するのは単純に過ぎる。有意でないデータから一般性につながる推論を導くことは不可能であるが、論文から有益なヒントが得られることがある。ただし、そうした場合には実験方法の再検討を含めた多角的な考察が必要になる。後で論じることにした。

統計的検定の新しい指標 日本で推測統計法が使われ始めた頃、研究者の多くは用例を「模倣して」計算を実施していたのではないかと思われる。つまり、岩原 (1965) の「教育と心理のための推計学」など、テキストに載っている多くの検定実施例から、自分のデータの構造に合致する事例を選び出し、その事例の演算の各ステップに自分のデータの対応する数値を当てはめ、得られた t 値や F 値などを、原理の理解をカッコに入れたまま、有意性検定の結果として報告することが多かったのではないかと思われる。その後、ほとんどの大学で心理統計法が必修科目になり、学部段階から統計法の系統的な教育が行われるようになった。また、基礎原理を丁寧に解説した良質のテキストも刊行されてきている。推測統計法は心理学研究の重要なツールとして本格的に活用されているといつてよい。ところで、最近になって、こうした推測統計法の活用に新たな展開が現れ、大方の支持を得てきている (Fritz, Morris, & Richler, 2012)。

これまでの統計的推測では、要因の効果がゼロであると仮定する「帰無仮説」null hypothesis にもとづく推測が主体であった。たとえば、群や条件の間に差があることを証明したい場合、群や条件の間に「差はない」とする帰無仮定をおき、その否定、つまり、否定の否定を通じて「差がないとはいえない」という推論を導いてきていた。こうしたアプローチについて、南風原 (2014, p.5) は、「検定統計量が統計的に有意か否かの 2 値的な判断をするのみで、データから得られる情報の活用という点で限界がある」こと、また、「帰無仮説が真である状況は想定しにくい」といった方法論

上の問題が多く指摘されていることを説明し、こうした検定の限界を補完する、または、代替する結果の提示法として、平均値差や相関や連関の実際の大きさ、すなわち、効果量を積極的に活用することが推奨されていると解説している。

アメリカ心理学会 APA はこうした動向に対応し「論文作成マニュアル」の改訂を実施してきている。21 年前の APA Publication Manual 4th Edition (American Psychological Association, 1994, p.15) では、 t テストや F テストやカイ自乗テストなどの推測統計法の結果を報告する際、テストで得られた値、自由度、確率水準、効果の方向などに加え、平均や中位数、それらの対応する変動の指標などの記述統計量を記載するように求めていた「だけ」であった。しかし、その後の Edition では新たな事項、効果量 effect size や信頼区間 confidence intervalなどを記載するよう内容が改められてきている。さらに、同じマニュアルの最新版 6th Edition (American Psychological Association, 2009, p.116) では、上記の諸テストや効果量や信頼区間などの推測統計量を報告する際には、行われた分析を読者が十分に理解できるよう、必要な情報の提示を求めている。すなわち、読者が論文に報告されている基本的な分析を確認できるようにするだけでなく、関心のある読者が、報告されている範囲を超えて、効果量や信頼区間の算定ができるように、細胞平均 cell means、標準偏差、標本の大きさ、相関値などのデータを提示することなどを求めている。

新たな指標の理解 上記のマニュアルには効果量や信頼区間などの新たな指標を表記する凡例が示されている (p.117)。分散分析の場合には、凡例のスタイルに沿って、従来からの検定指標に「加え」て、効果量と信頼区間を示すやり方が定着しつつあるようである。具体例を示そう。たまたま手許にあった Gilroy & Pearce (2014) の論文で、ラットの「弁別比」を「群」と「テスト文脈」の 2 要因で分析した分散分析の結果である。

群の要因は有意であった。 $F(1, 34) = 6.58$, $MSE = .043$, $\eta_p^2 = .16$, 95% $CI [.01, .37]$ 。
しかし、テスト文脈、および、群とテスト文脈の交互作用は有意ではなかった。

η_p^2 以降が新たな項目である。 η_p^2 は効果量の指標の 1 つである偏決定係数、95% $CI [.01, .37]$ は効果量の 95% 水準での信頼区間である。偏決定係数はその要

因の持つ効果の大きさ、正確には、その要因をモデルに含めることによって、その要因を含める前に説明できていなかった残差分散の何%を説明できたかを示している(南風原, 2014, p.83)。偏決定係数は効果量の点推定値である。一方、信頼区間は「結果と整合的な母数の値の範囲」を表している。この事例では、母集団における決定係数(分散説明率)がわずか1%の状態から37%の説明率までの広い範囲に及びうことを表しており、この広い範囲の可能性を勘案して分析結果を理解する必要があることを示している。効果量と信頼区間は、 t 検定やカイ自乗検定などについても報告することが推奨されている。

新たな統計指標の追加 以前の論文には、効果量や信頼区間の指標は記載されていない。しかし、こうした論文についても、然るべきパラメータが入手できれば、読み手の側で効果量や信頼区間を計算することができる。そして、その結果にもとづいて、要因の効果の母数の広さ・狭さを考慮に入れた評価を「追加的に」行うことができる。分散分析の変動因の効果量は分析のモデルと各要因の自由度が分かれば計算できる。たとえば、Fritz et al. (2012)や南風原(2014)にそれに関する説明がある。一方、信頼区間の計算は手計算では困難であると思われる。南風原(2014)は F 値の上側下側確率がそれぞれ $\alpha/2$ ⁴⁾となる非心度パラメータ λ を統計パッケージ「R」で求め、信頼区間の限界値を計算する手順を説明している。一方、Steiger (2004, p.168)は、 F テストと t 検定とカイ自乗テストについて、信頼限界値の λ を計算することのできる無償のソフトウェア“NDC”を自らのホームページで公開している。

「考察」を検討する

論文評価の多元性 論文の「方法」を検討している段階で、早々に論文の評価が定まってしまう場合がある。端的な例は、実験計画が必要な統制条件を欠いている論文である。これほど直裁的ではないものの、データ収集の方法に疑念があり、実験結果の信頼性を検討する前に、評価が否定的になってしまう場合もある。顕著な例は、Garcia 効果を報告した初期の論文の1つ(Garcia & Koelling, 1966)である。Garcia 効果とは、条件づけがCSとUSとの組み合わせに依存することを指し、「生物学的制約」の証拠として広汎な関心を集めた現象である。Garcia らはCS-USの組み

合わせ3種類4条件についてのデータを示し、内部的CSは内部的有害刺激と、外部的CSは外部的有害刺激と結びつきやすいと主張する論文を発表した。しかし、この報告は別々に実施した実験結果をまとめて比較した証拠にもとづいており(小牧, 2012, p.63-66)、通常の実験の条件間比較の要件を満たしていなかった。考察にはこの点についての説得的な分析はなかったことから、この論文は「探索的な実験報告」として評価するほかはなかった。幸い、Garcia らは、それとは別に、同一実験の中で条件を比較する研究を報告しており(Garcia, Ervin, & Koelling, 1966)、上記のCSとUSとの交互作用についてはじめて確かな証明を提示したことになる。

実験結果の信頼性を検討する段階で、懐疑的な評価を下さざるを得ない論文もある。統計処理に疑問がある論文はその1例である。さらに、統計的な有意性が証明できなかった報告も、もう1つの例である。有意性の意味をめぐってはさまざまな議論があるが、有意性の判定は実験結果の学問的価値とは次元を異にする、形式的な基準にすぎないことを指摘しておきたい。実験結果が有意でなかった論文については、先に述べたように、一般につながる推論を導くことは困難であり、考察の扱いには慎重さが求められるが、一切の検討を放棄するのは早急に過ぎる。そんな論文でも、考察や実験結果に、今後の研究に役立つ示唆的な情報が含まれている可能性がある。方法に格別な不備が発見できなかった実験は殊にそうである。こうした報告については、標本の標準偏差や分散など、データの変動を点検することを推奨したい。たとえば、類似の実験に比べ、条件ごとの個体間変動が大きい場合には、実験手続きの不備と将来の改良の必要性が示唆される。一方、被験体の数が少ないにもかかわらず、個体間分散が類似の実験と同等、または、小さい場合には、個体数を増やすことで、要因効果が有意となり、信頼性の確保される可能性が見込まれる。こんな場合、どこまで個体数を増やすべきか、それが問題になるが⁵⁾、私は「個体数の相場」(小牧, 1992, p.150)が1つの目安になると考えている。ただし、最近の論文には、標本の標準偏差や分散を報告しているものは少ない。そん

5)：大久保・岡田(2012)は、検定力、効果量、有意水準、および、標本サイズの関連性について論じており、注目される。ただし、実験は効果量そのものが不明であることから行われることが多い。この意味で、効果量の事前予測にもとづいて、適正な個体数を定めることはむずかしい。次章で改めて論じることになしたい。

4)： α は F 値の有意水準。

な時には、 F 検定の MSE などで代替して判断するわけではない。

「目的」の種別と達成度の評価 信頼性を確認した実験結果が、読み手の予測と格別な矛盾がない場合、読み手はその結果が実験目的の確かな回答になっているかどうかを分析的に点検し、論文の考察の評価に取りかかることになる。この際、論文の冒頭に戻り、実験の目的を改めて確認しておいた方がよい。実験は目的に応じて「探索実験」「確認実験」「批判実験」の3種に大別できる(小牧, 2000, p.24-34)が、目的に応じて、考察評価のポイントにも違いが出てくる。

「探索実験」は、研究を進める手がかりとなる証拠を手に入れるため、手探りにやってみる実験であり、形式的な完結性を満たしていないケースが多い。しかし、探索実験に求められるのは形式よりも内容である。その実験がどのような想定にもとづいて計画されたのか、その想定は支持されたのか、実験の結果から研究の方向を定めるためのどのような証拠が得られたのか、こうした点が探索実験を評価するポイントになる。探索実験の「考察」が単なる結果の記述に終始したのでは物足りない。実験結果を踏まえた将来の研究の見通しやかんばしい結果が得られなかった場合の見込み違いの原因究明など、視野の広い論証が期待される。そして、何よりも重要なのは、将来の研究の展開についての著者の見通しである。研究者の問題関心やアイデアは、仮説または理論の形に結実してはじめて検証が可能になることを重視したい。

「確認実験」は、基本的には研究者の考えを検証するための実験であるが、検証する考えの領域と水準は多様である。ちなみに、実験変数の一般性を確かめる拡張実験、理論的予測の是非を確かめる仮説検証実験など、心理学実験の多くはこのタイプである。確認実験の最大のポイントは、実験計画が検証する問題にふさわしい設定になっているかどうかにある。たとえば、理論的な予測を検証する仮説検証実験は、方法が問題の過程を適正に生起させていなければ、まったく意味を持たない。見当違いの「空振り」(小牧, 2000, p.15)の実験になる。ほかに、必要な統制条件が設置されているか、適切な測度が用いられているかなど、こうした問題についての著者の主張を読み取り、実験計画の「妥当性」を確認することが確認実験評価の、最も重要なポイントになる。

「批判実験」は確認実験の特殊なタイプであり、対立する理論のうちのどの理論が正しいのか、決着をつ

けることを目的とする実験である。対立する理論を1つの実験の中で対峙させるためには、個々の理論を正確に理解し、実験計画に合わせて正しく拡張することが実験の要件になる。この要件の確認には、読み手の側が問題の理論を正確に理解していることが前提になる。批判実験は心理学実験のうちの最もスマートなタイプであるが、関連理論の基本的な対立点に適正に焦点を当てることでこのスマートさが生かされることに留意したい。

予測を覆す新たな知見 方法を読んだ段階で、果たして想定されているような結果が得られるものか確かな予測ができなかった実験がある。しかもそうした実験で、それまでの常識を覆すような発見が得られた報告がいくつもある。その中の最も教訓的な最近の事例は「知覚学習」に関する Channell & Hall (1981) の実験ではないかと私は考えている。

知覚学習とは、Gibson & Walk (1956) が報告した「形の弁別」に関する学習作用で、ラットに正三角形と円の刺激図形を観察する機会を与えると、それだけで刺激の弁別学習が促進されるという作用を指す(小牧, 2012, p.202-204)。この観察には強化が随伴していなかったことから、Gibson らは、この促進作用は強化なしにラットが「刺激の微細な違い」に注意するようになったことを示唆する証拠であると考えた。ところが、Hall らは、この無強化刺激提示には「古典的条件づけ」の「潜在制止」と同質の、妨害的な作用が生じるはずだと推測していた。つまり、刺激の無強化提示は、その刺激の「連合可能性」associability を低下させ、その後の学習を妨害するはずであり刺激提示が本来的に促進的な作用を持つはずがないと考えていたのである。

Hall らは、この逆説的な作用について、「学習環境の変化」が関係していると見こんでいた。これには傍証があった。潜在制止は環境変化に依存することが知られており、変化しただけでは制止効果が逆転する場合があることが報告されていたのである(Channell & Hall, 1981, p.438)。ちなみに、Gibson らの実験では、刺激提示は飼育箱でテストの弁別学習は「グライス箱」でと、訓練とテストで環境が大きく変化していた。Hall らはこのことに着目し、「変化のある」条件と「変化のない」条件とを設置した実験を計画した。すなわち、飼育箱で刺激を提示し、「ジャンピング・スタンド」で弁別を訓練する、Gibson-Walk 実験と同様の転移条件とジャンピング・スタンドで刺激提示と訓練との

「双方」を行う転移条件とを設けたのである。これらの2種の実験群条件と対比するため、Hallらは飼育箱で刺激を提示しない統制条件を加えている。

実験結果はHallらの予測を支持した。変化あり条件ではGibsonらと同じ促進効果が現れた。これに対し、変化なし条件では、推測したとおり妨害効果が現れたのである。刺激の無強化提示が学習妨害的な抑制作用を持つとして、その作用が環境変化との交互作用を介して促進作用に転じるとするHallらの予測は、少なくとも私には、ありそうもないことに思われた。しかし、実験の結果は、彼らの予測を支持した。私の予測の誤りは、潜在抑制が刺激の独自要素と共通要素とに分化的に作用するという理論的可能性に思い至らなかったことであった。つまり、刺激のそれぞれは固有の独自要素と他刺激と共通する共通要素とから構成されているが、共通要素は抑制を受ける機会が独自要素の2倍ある。このことから提示されている刺激の共通要素はより大きな抑制をこうむり(小牧, 2012, p.204-211), 結果として個々の刺激の独自性の強調, つまり、刺激の分化が進捗するという理論的可能性のことである。

Hallらの結果はその後の実験においても追証された。また、彼らの潜在抑制に基礎を置く知覚学習理論は、「放射状迷路」や「味覚条件づけ」などの事態にも適用され検証を得ている。ちなみに、Mackintoshはその遺作となった知覚学習の評論(Mackintosh, 2009)において、Gibson-Walk効果を「分化した潜在抑制」*differential latent inhibition*の事例として説明する理論的展望を提示している。

知識の蓄積と研究者 Gibsonらの知覚学習の論文は1956年に発表された。潜在抑制に関するLubowらの最初の報告(Lubow & Moore, 1959)はその3年後に発表されている。ともに無強化刺激提示に関わった実験報告であり、両報告の類縁性に気づいた研究者は少なからずいたと思われる。しかし、一方は促進作用であり、もう一方は抑制作用であった。1976年のLubowらの潜在抑制の環境依存性の指摘(Lubow, Rifkin, & Alek, 1976)はあったものの、両作用を統一的に説明する視点の展開は、1981年のHallらの実験的証明を待たねばならなかった。この経緯は、新たな理論の構築には、研究者の積極的な文献検索に加えて「思索」が必須であることを明確に語っている。図書館の書庫は理論を形成してはくれない。また、パソコンのハードディスクは実験仮説を提起してはくれないのである。

新たな構想を得るためには、理論からの示唆、同学の研究者との討論など、いくつもの資源の活用が必要だと思われる。しかし、研究者個人の思索の展開には、論文の読みこなしを通じた法則的知識の蓄積が不可欠である。「こうすれば」「こうなる」という法則的知識の集合は、未整理の知識の混在する体系化の不完全な知識集合に過ぎない。しかし、この知識体は、「こうすれば」という手続き言語と「こうなる」というデータ言語で表される経験的法則の集合であり、理論とは「距離を置いた」情報源である。この意味で、個々の研究者の法則的知識の総体は、その研究者の諸知見に対する評価の基礎となるとともに、理論批判と理論改訂の基礎として機能すると私は考えている。

論文は教科書ではない。そのままを受動的に受け入れるのは、論文の本来的な読み方ではない。論文は、研究者が、自らの法則的知識を土台にそれと対峙し、批判的にかかわって行くべき検討対象である。その内容に同意できないことがある。その主張に納得し、自らの知識体系を修正せざるを得なくなることもある。そのいずれもが論文との積極的なかわりのしるしであり、研究者の能力発達のあかしになるのではあるまいか。

引用文献

- American Psychological Association (1994). *Publication Manual of the American Psychological Association*. 4th ed. Washington DC: American Psychological Association.
- American Psychological Association (2009). *Publication Manual of the American Psychological Association*. 6th ed. Washington DC: American Psychological Association.
- Channell, S., & Hall, G. (1981). Facilitation and retardation of discrimination learning after exposure to the stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, **7**(4), 437–446.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, **141**(1), 2–18.
- Garcia, J., Ervin, F. R., & Koelling, R. A. (1966).

- Learning with prolonged delay of reinforcement. *Psychonomic Science*, **5**(3), 121–122.
- Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, **4**(1), 123–124.
- Gibson, E. J., & Walk, R. D. (1956). The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *Journal of Comparative and Physiological Psychology*, **49**(3), 239–242.
- Gilroy, K. E., & Pearce, J. M. (2014). The role of local, distal, and global information in latent spatial learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, **40**(2), 212–224.
- 南風原朝和 (2014). 統・心理統計学の基礎－統合的理解を広げ深める 有斐閣
- 岩原信九郎 (1965). 教育と心理のための推計学 (改訂版) 日本文化科学社
- 小牧純爾 (1992). 変動と一般性：個体型実験法の理解をめぐって 心理学評論, **35**(2), 133–155.
- 小牧純爾 (1995). データ分析法要説－分散分析法を中心に ナカニシヤ出版
- 小牧純爾 (2000). 心理学実験の理論と計画 ナカニシヤ出版
- 小牧純爾 (2012). 学習理論の生成と展開 ナカニシヤ出版
- Lubow, R. E., & Moore, A. U. (1959). Latent inhibition: The effect of nonreinforced pre-exposure to the conditioned stimulus. *Journal of Comparative and Physiological Psychology*, **52**(4), 415–419.
- Lubow, R. E., Rifkin, B., & Alek, M. (1976). The context effect: The relationship between stimulus preexposure and environmental preexposure determines subsequent learning. *Journal of Experimental Psychology: Animal Behavior Processes*, **2**(1), 38–47.
- Mackintosh, N. J. (2009). Varieties of perceptual learning. *Learning & Behavior*, **37**(2), 119–125.
- 大久保街亜・岡田謙介 (2012). 伝えるための心理統計－効果量・信頼区間・検定力 勁草書房
- Steiger, J. H. (2004). Beyond the *F* test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, **9**(2), 164–182.

(2015 年 4 月 24 日受稿, 7 月 3 日受理)